Xiangyu Jin
2135617608 • xjin8592@usc.edu

University of Southern California
Master of Computer Science *Aug 2024 – 2026(expected)*

National University of Singapore
Master of Science in Data Science and Machine Learning (GPA:3.78) *Aug 2023 – July 2024*

University of Illinois at Urbana Champaign
Bachelor of Science in Statistics (GPA:3.76) *Aug 2019 – May 2023*

PERSONAL EXPERIENCE

Scenes Image Classification *Mar 2024 – May 2024*

Group Leader

- Applied pre-processing techniques on image data; designed CNN architecture to train the model to classify images from 15 different categories; achieved 72.44% accuracy on test dataset.
- Utilized pretrained weights Resnet50 to fit the model and resulted in 98.64% accuracy.

House Price Prediction Competition (Kaggle) *Jan 2024 – May 2024*

Individual

- Performed data cleaning, preprocessing, and feature engineering using python.
- Trained Lasso, Ridge, Random Forest, XG Boost, and Neural Networks to compare performance. Used model ensembling to achieve the highest accuracy. Achieved RMSE 0.12438 and ranked top 10% in the leaderboard.

Image Compression with Mixed Transformer-CNN *Feb 2024 – May 2024*

Group Member

- Realized the method described in paper, Learned Image Compression with Mixed Transformer-CNN Architectures(arXiv:2303.14978v1).
- Built CNN-Autoencoder model and GANs model in Python to compare compression efficiency; replaced dataset to test generalization of the models.
- Operated Transformer-based models and mixed CNN-based layers in the architecture.

Movie Database Project *Oct 2023 – Nov 2023*

Group Leader

- Combined data from IMDb and TMDB to build a relational database using MySQL and a document database using MongoDB, cooperating with TMDB API. Compared complexity of building relational database and document database and efficiency of querying in different databases.
- Proficient in building databases and querying data using MySQL and MongoDB.

Shiny Dashboard Project for Music Analytics *Mar 2023 – May 2023*

Group Leader

- Developed an interactive Shiny Dashboard that allows users to explore artist information, access an artist's top 10 songs, and visualize how an artist's popularity has evolved over the past decade.
- Proficient in utilizing the Spotify API in R programing language for music data retrieval, including artist information and album details.
- Implemented machine learning models, including multinomial logistic regression, support vector machine, and XGBoost to predict singer hotness trend.

INTEREST AND DISTINCTION

- Computer: Python, R, MySQL, Java, C++, Excel, PowerPoints
- Language: Mandarin (Native), English (Fluent)
- Interests: Badmintons, Movies, Cooking, Dog Training